

DOCUMENT RESUME

ED 342 212

FL 019 255

AUTHOR Ingram, D. E.
TITLE Language Proficiency Testing in the International Setting: Bibliographies, Portfolio of Papers, and an Overview Paper.
PUB DATE Jun 90
NOTE 46p.; Paper presented at a meeting of the International Working Group on Assessment of Proficiency Levels for Students of Japanese as a Foreign Language (Canberra, Australia, June 12-14, 1990).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Reference Materials - Bibliographies (131) -- Speeches/Conference Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Annotated Bibliographies; Elementary Secondary Education; Foreign Countries; *Japanese; *Language Proficiency; *Language Tests; *Rating Scales; Second Language Instruction; *Second Languages; Testing; Uncommonly Taught Languages
IDENTIFIERS *Australia

ABSTRACT

Materials compiled for a conference on the assessment of Japanese second language proficiency include a partially annotated bibliography of references concerning proficiency testing as it may relate to the Japanese language and an overview essay entitled "Towards the Development of Proficiency and Other Tests in Japanese as a Foreign Language in Australia." The essay contains a review of literature concerning the nature of language proficiency and its assessment, a historical review of approaches to assessing proficiency, a discussion of the needs that Japanese tests and scales must meet in Australia, the types of tests required to meet those needs, and notes on available tests. A concluding section looks at directions for future proficiency testing of Japanese as a second language. A wide range of needs related to proficiency testing, and their implications for test design, are outlined, including the needs for: a battery of tests for a variety of purposes; proficiency testing alternatives; industry- and commerce-related language tests; language program development; educational quality control; diversification of and increase in language learning; testing for achievement, placement, and diagnosis; self-assessment methods; classification of materials and activities; development of program rationales; testing of aptitude, cross-cultural attitudes, and cultural understanding; research data; and examiner training. (MSE)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED342212

LANGUAGE PROFICIENCY TESTING IN THE INTERNATIONAL SETTING:
BIBLIOGRAPHIES, PORTFOLIO OF PAPERS, AND AN OVERVIEW PAPER

OVERVIEW PAPER: Towards The Development of Proficiency and
Other Tests in Japanese as a Foreign Language in Australia

D. E. Ingram
Professor of Applied Linguistics and Director,
Centre for Applied Linguistics and Languages,
Griffith University, Brisbane, Australia

Prepared for the Meeting of the International Working Group on Assessment
of Proficiency Levels for Students of Japanese as a Foreign Language,
Canberra, 12-14 June, 1990.

PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Ingram, D.E.

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

FL019255

2

OVERVIEW PAPER: Towards The Development of Proficiency and
Other Tests in Japanese as a Foreign Language in Australia

D. E. Ingram

CONTENTS

- I Preamble
- II The Nature of Proficiency and its Assessment
- III Assessment of Proficiency
- IV Needs that Japanese Tests and Scales Must Meet in Australia
 - Battery
 - Proficiency
 - Industry
 - Language Programmes
 - Quality Control
 - Diversifying and Increasing Language Learning
 - Attainment Tests
 - Streaming
 - Diagnosis
 - Self-Assessment
 - Classifying Materials and Activities
 - Programme Rationalisation
 - Other Tests
 - Research
 - Training in Testing
- V Test Types
- VI Tests of Japanese
- VII The Future

[NB. References in this paper can be found either in one of the bibliographies or in the reference list for "Assessing Proficiency" in the portfolio of papers.]

3

I PREAMBLE: In preparation for this conference, I was asked to provide an overview of language proficiency assessment in recent years and, specifically, to provide an annotated bibliography of significant works, an unannotated bibliography, and a background paper which would review the literature. I must say that I approached the task with some trepidation since, through the 1980's, the literature on proficiency and its assessment has proliferated. Indeed few concepts in applied linguistics have generated so much scepticism when first explicitly introduced and subsequently so much interest and activity. The scepticism, which is not entirely in the past, is difficult to explain unless one recognizes the seductive hold that the false objectivity and false security of statistics and psychometrics have exerted over language testing. The reasons for the interest and the activity are more readily explained since proficiency is at the very heart of what language teaching is about: whatever else language teachers aspire to do, they must seek the development of proficiency, that is, they must seek to develop in learners the ability to use the language for practical communication tasks. In fact, virtually everything else that language teachers have both traditionally and more recently sought in language teaching (literary appreciation, aesthetic response, cultural understanding, translation skills, ability to use the language in everyday or in specialized situations, and so on), all are dependent on the learners' developing the ability to use the language. To be fair to the sceptics, it has also to be said that some of their scepticism arises from academics' legitimate fear of things that cannot be unambiguously defined: though everyone, both applied linguists and laypersons, uses the term "proficiency" and knows what it means when they say that "Bill is proficient in French" or "Mary is proficient in Japanese", it is in fact difficult to arrive at a fully explicit definition of the concept and this seems to make the development of tests to measure it more difficult and more open to attack.

The other reason why I approached with trepidation the task of preparing a review paper for this conference was that, having prepared quite lengthy bibliographies, I felt that the review paper might become somewhat repetitious. In addition, I had already some years ago written a lengthy paper looking at the different types of tests and assessing their usefulness for the measurement of proficiency and, more recently, Elaine Wylie and I had written a paper in which we reviewed the different approaches that could be adopted in developing proficiency scales: both of these papers are in the portfolio. Finally, I was also conscious of the fact that the audience to whom I would be addressing this paper was highly select, being either laypersons who are end-users of language skills and test results or language testers who could be expected to be very familiar with the concept of proficiency, its assessment and the associated literature.

Consequently, in preparing for this conference, I have sought to survey the literature as comprehensively as possible, select a cross-section of items for inclusion in the annotated and unannotated bibliographies, and survey whatever tests of Japanese proficiency are already available; I have tried to provide in the portfolio, papers which were immediately relevant to the task of test development, which would exemplify some of the different proficiency instruments available and raise some key issues to be considered by persons charged with developing tests. Consequently, this

paper does not attempt to review the literature (the bibliographies do that) but it aims to provide an overview of the issues that are relevant to the purpose of this meeting and of the working party, to point to the directions in which our thoughts and discussions might run, and to suggest the broad issues to be considered as we seek to devise ways by which the field of language testing and of proficiency assessment in particular might respond to the need for Japanese language skills in Australia both now and into the 21st Century.

II THE NATURE OF PROFICIENCY AND ITS ASSESSMENT: In the paper on "Assessing Proficiency" in the portfolio, it was observed that, over the last forty years, the concept of proficiency and the approach to measuring it have changed as our understanding of the nature of language and of language learning has changed. When language was seen as a set of rules and vocabulary that were memorized and consciously applied, proficiency in a language was equated with knowing the grammatical rules and lists of vocabulary and being able to formally apply them. From here, "proficiency" passed through behaviourist notions of an accretion of conditioned habits, then the internalized rules of cognitivism, until, as functional-notional and communicative ideas developed, applied linguists and testers departed from the analytic approach, realized that language is more than the sum of its parts, and moved the focus onto the tasks that learners can carry out in the language and how they are carried out. In other words, having passed through the "traditional", the behaviourist, the cognitivist, the functional-notional and the communicative stages, the definition of proficiency now is seen to embrace both the learners' ability to carry out tasks in the language and how those tasks are carried out; rather than being analytic, the approach is now synthetic and global, with the focus on the learners' total language behaviour or performance. Characteristically this behaviour is described in scales which seek to capture the sort of language behaviour or performance that can be observed in learners as their language develops from zero through to native-like. This also means that, whereas more traditional approaches to testing are analytic in that they seek to analyse the language into its component parts and measure proficiency by adding up the learners' scores on each of the discrete bits of the language, the overriding though not exclusive emphasis today in assessing proficiency is on the learners' overall or global performance with their knowledge of individual grammatical rules, vocabulary items or other elements of the language primarily of interest insofar as they contribute to overall language behaviour.

There have been many attempts to define proficiency and to capture its principal features. Some of these are referred to in the paper "Assessing Proficiency". Vollmer [1981] seems to opt out and defines proficiency as what proficiency tests measure. Politzer and McGroarty [1983] define it simplistically as the ability to convey or receive information, which seems to omit consideration of the way in which the information is conveyed but, in fact, they claim to show in their study that, though communicative competence is distinct from linguistic competence and the learner may have strategies that enable him or her to compensate for deficiencies in linguistic competence, linguistic competence nevertheless imposes limits on communicative competence. Perhaps the most cited characterisation of proficiency is that of Canale and Swain [1980] in which they identify four components, viz., grammatical competence (in the widest sense of "grammar" to entail all the rules by which utterances are put together), sociolinguistic competence (or knowledge of the sociocultural rules of use and of discourse), and strategic competence (or the set of verbal and non-verbal communication strategies that learners adopt "to compensate for

breakdowns in communication due to performance variables or to insufficient competence") [Canale and Swain 1980: 30; Swain 1984: 9]. Bachman [1988] defines proficiency as the ability to conduct culturally-embedded verbal interaction with speakers of the target language and he identifies a similar array of underlying features as do Canale and Swain with, in addition, reference to motor skills. An important distinction made by Griffin and others [1988] and made also in the literature on the ASLPR [e.g., Ingram 1979/1982, 1984 in unannotated bibliography, 1985] is that between course-dependent performance using the material taught during the course and general proficiency or, in the terms used by Griffin et al, "the effective use of language in contexts...irrespective of the body of material taught in class" [1988:1]. A further essential distinction is made in "Assessing Proficiency" between proficiency and "communicative competence" in the loose sense of the ability to communicate: the ability to communicate and the success of attempts at communication depend on things beyond facility in the language itself including such factors as intelligence, knowledge and the ability to organize it, introversion or extroversion, and the willingness of the interlocutor to accommodate less than native-like language. Proficiency (at least in the sense in which it is used in the documents on the ASLPR [Ingram 1979/82, 1985]) relates to facility in the language as it is used to carry out language tasks.

III ASSESSMENT OF PROFICIENCY: Approaches to testing or the assessment of proficiency have also differed according to the view of language and its learning dominant at the time. The analytic approaches referred to above, described in "Assessing Proficiency" (in Sections II.2 and III.2), and characteristic of so-called indirect tests assessed proficiency by analysing the language into its component parts and assessing the learners' knowledge of or ability to use those component parts. Although such approaches are very effective for diagnostic testing, in assessing proficiency the problem is, first, that language is more than the sum of its component parts and, second, the sum of the items scored correctly on a test does not necessarily give much indication, or any readily interpretable indication, of how well the learners can use the language in non-test, real-life situations.

Predominantly in the late-1960's and the 1970's, there was much interest (illustrated especially in Oller's work [e.g., 1979, 1983]) in developing test types that focus on discrete items in the language but do so in a total language context. These tests, which drew on the redundancy of language and, in Oller's terms [1973, 1973a, 1974], on the learner's expectancy grammar, sought to be as objective as indirect tests, to focus on individual items, but to embed them in a full text. Characteristic of these tests are dictation, white noise, interlinear and the various types of cloze but, as with indirect tests, their interpretation is difficult and, however effective they might be for some diagnostic purposes and for rank-ordering learners, it is difficult to move from the scores to an unambiguous statement about what the learners can do in the language.

To overcome problems of interpretation and to try to assess proficiency in situations where the language is operating more naturally (i.e. with all its elements contributing to carrying out language tasks, to achieving some purpose, and in a sociocultural context), approaches have been sought that allow the language to operate naturally in a communication situation in which the learner is carrying out normal language tasks. The assessment process is a matter of rating the observed language behaviour against a scale which describes learners' language behaviour as it develops from zero to native-like. This is commonly done by observing the learners' use of the

language in an interview designed for the purpose, during which the observed behaviour is matched against the scale descriptions. However, other approaches are also possible. Some testers [e.g., Brindley 1986, Pieneman et al 1988] see a close relationship between proficiency levels and the emergence of certain elements of the language (especially syntax) and devise tests that focus on these features in order to relate learners to proficiency levels. Rating using proficiency scales is essentially a process of observing language behaviour and matching it against scale descriptions and this observation or, more accurately, the process of eliciting language behaviour to observe, may be encapsulated in more traditional types of tests with either the scores or performance on individual items related in some way to the scale levels. Ellis [1984] seems to abandon the attempt to identify the nature of proficiency, theorizes that the greater the learners' proficiency the fewer "communication strategies" they will use, and attempts to use this phenomenon to rate learners' proficiency. Others [e.g., Politzer and McGroarty 1983] use an analytic approach to rating language performance by specifying in considerable detail the elements of language to be observed and how they are to be assessed. The newly developed International English Language Testing System uses different approaches to relate language performance to descriptive scales for each macroskill: the Speaking test uses direct matching of behaviour and global descriptors, the Writing test uses more analytic sub-scales converted statistically to an overall Band Scale, and the Listening and Reading tests statistically convert raw scores on comprehension tests to Band Scale scores [see Ingram 1990, 1990a].

Tests designed to assess some notion of proficiency not only differ on the "directness" with which they approach the task of assessing proficiency but also on the basis of construction of the test. In "Assessing Proficiency", a distinction is made between non-developmental and developmental tests. In non-developmental tests, proficiency is viewed more or less statically as facility in handling certain language content which may be chosen for inclusion in the test on an ad hoc, a linguistic, or a behavioural basis. In contrast, developmental tests see proficiency as developing in an identifiable way from zero to native-like ability levels and they are designed to identify at what point along the continuum from zero to native-like the learner's proficiency falls. Proficiency scales such as the ASLPR, the IELTS Band Scales, the ACTFL Guidelines, the FSI Scale, and the ILR Scale all essentially assert that it is possible to describe how proficiency develops and to slot learners into this continuum. This approach has particular advantages since, as in the ASLPR for example, proficiency is assessed by identifying the sort of language behaviour that the learner manifests, the tasks that can be carried out and how they are carried out, and the point on the scale that most nearly matches the observed behaviour; but such an assessment also indicates the point to which the learner's language has developed and, hence, the direction of future development. Clearly this has important implications for language teaching and enables proficiency scales to make a significant contribution to the teaching process, especially in syllabus design, methodology and the selection of materials and activities [see Ingram 1986].

It was noted earlier that the concept of proficiency lies outside any particular course that learners might follow. The last point just made, however, also suggests that the concept of proficiency is not language specific, i.e., the literature of developmental psycholinguistics clearly suggests that the broad schedule of development is universal even though details of development, the precise ordering of some syntactic forms, for instance, might differ to some extent. Intuitively, we recognize this since

we have at least some vague notion of what we mean if we say that Martha is proficient in Thai and Tom has just enough Motu to survive in the village. It is also suggested in practice by the fact that scales such as the FSI, the ILR, ASLPR and ACTFL are successfully used to rate the proficiency of learners of any language. This observation is important, however, because it makes the development of proficiency scales for the wide range of languages necessary in Australia more practical since it suggests that one does not need a proliferation of scales but rather a common, broad scale which may also include those features distinctive to particular languages. Thus, for instance, the ASLPR is used to rate proficiency in any language, it has been formally assessed and reviewed in the context of ESL, French, Italian and Japanese, the General Description column is common to all versions, and an example column is used to illustrate how those general descriptions relate to particular languages [see Ingram and Wylie 1979/1985 and the versions for French, Italian and Japanese]. For the immediate purposes, this also means that any scales developed for the assessment of Japanese proficiency will probably carry over significantly to other languages and whatever other tests are developed may also serve as useful paradigms for other languages.

In this very brief description of the different sorts of tests that occur and their basis of construction, it is clear, I hope, (and is made clear in "Assessing Proficiency") that different test types are essentially more or less effective, and hence more or less appropriate, for different purposes. Indeed, if one puts aside the possibility of errors in test construction, it is correct to say that there is no intrinsically good or bad test type but rather different test types are more or less effective in achieving different purposes in different contexts; in order to select a test or test type, one must consider the purpose for which one is testing and the context or constraining factors in the test situation and match the test or test type with the purpose and context.

Consequently, in considering what sorts of tests of Japanese may be desirable in Australia, it is essential to consider the purposes for which one requires tests, the needs that testing has to meet, and the contexts in which testing has to take place.

IV NEEDS THAT JAPANESE TESTS AND SCALES MUST MEET IN AUSTRALIA:

The needs for Japanese tests in Australia are wide-ranging and, when test construction is to be undertaken, should be considered in some detail. Here we shall only attempt to get an overview of the wide range of needs that exist and the implications they have for test design. In the next three days, these and other needs should be clarified and the test implications elaborated.

Battery: It is already evident that there is a variety of purposes which tests of Japanese have to serve and the different needs will require different test types. It may well be, therefore, that the outcome of this present project should not be just a new test or proficiency scale but proposals for a battery of tests for use with learners of Japanese. How many of the needs can be met and the tests or scales developed will depend on the resources available.

Proficiency: The preceding discussion suggests that, in order to assess proficiency, a direct assessment procedure is desirable but, in fact, this will not always be possible. The initial need is undoubtedly for a

proficiency scale, for interview and other elicitation procedures to be worked out, and for alternatives to interviewing to be proposed where it is not possible to use a fully trained native speaking interviewer. In other words, there is need for a proficiency scale or scales in Japanese to be developed but also for the contexts in which it or they will be used to be considered and a range of application methods to be developed. It may be, for example, that, in addition to standard interview procedures, it would be useful to consider the approach adopted by the International English Language Testing System (IELTS) which uses a wide range of test types and includes a controlled Speaking interview designed to be used by minimally trained teachers; it is possible that the IELTS provides a model that could be adapted for use with Japanese proficiency scales by non-native speaking teachers of Japanese. Two papers describing the IELTS approach have been included in the portfolio of papers for this conference. For Speaking also, it might be desirable to consider whether a "simulated oral proficiency interview" would be useful partly because of the relatively low proficiency levels of Australian teachers of Japanese and because persons outside the education system who need Japanese may have little access to adequately proficient and trained assessors. A short paper on SOPI's by Charles Stansfield has been included in the portfolio.

Whatever other tests are needed and whatever the approach adopted, a proficiency scale or scales should be considered an essential means of explicating and giving coherence to other tests as well as for other, especially course-related, reasons to be discussed subsequently. One of the most significant advantages of the development and widespread acceptance of a proficiency rating scale is that it provides a way to speak in concrete terms about the abstract concept of proficiency [cf. Bachman and Savignon 1986: 380]: it provides common terminology and a means by which teachers, syllabus writers, employers and others can speak more comprehendingly of proficiency rather than by using such vague and quite meaningless terms as "beginners", "average" or "advanced".

Any scale or scales developed will need to take account of the four macroskills and the need to have a different sub-scale for each macroskill. The basic scales would seek to describe general proficiency, they would be developmentally-based, and they would provide descriptors of language behaviour in each macroskill in Japanese as proficiency develops from zero to native-like. Development of suitable scales in Reading and Writing will be especially difficult (but no less desirable) because of the complexities of the Japanese orthography.

In addition, there is need for "special purpose" versions of any Japanese scales developed to facilitate assessment in special purpose contexts, e.g., in Japanese for airline booking clerks, Japanese for scuba-diver trainers, Japanese for tour guides, Japanese for marketing managers, and so on. Desirably these scales would be similar to the general proficiency scale but with adaptations that exemplify the scale in the particular special purpose contexts. The special purpose versions of the ASLPR currently being developed may illustrate what is required [cf. also Biersteker 1986, Doyle 1988].

One "special purpose" proficiency scale of particular importance is Japanese for Academic Purposes for use with students going to study in Japan. At present the only test available is the seriously inadequate Japanese Language Proficiency Test which, it has to be said, is unlikely to make any useful assessment of a foreign language learner's practical proficiency in Japanese or his or her ability to cope at a Japanese

institution. In developing a suitable test for use with intending students in Japan, it may be that this conference should consider whether or not an initiative comparable to that which led to the development of the International English Language Testing System is not possible [see Ingram 1990]. The IELTS probably provides a useful model of a test which is designed to provide information on candidates' general proficiency and their "academic purpose" proficiency and is designed to be administered worldwide in situations where there is minimal control over the quality of administrators.

It should also be noted that the need for proficiency scales extends throughout the education system and beyond it. The problem of providing proficiency scales for use where young children are learning a language is particularly difficult and fraught with theoretical and practical problems. The most difficult problem yet to receive adequate research attention is whether or not a single scale can serve for children and adults alike or whether, because scales essentially provide levels within the developmental span from zero to native-like, it is necessary to have one or more separate scales for children. However, as was the case at the end of 1989 when the national curriculum projects in Japanese, Chinese and Indonesian were reviewed, there is an almost desperate need for proficiency scales suitable for use with children throughout the education system in order to assess their proficiency in the language, to explicate the national curricula, to match activities and materials to the learners' levels, and, in many instances, to enable the national curricula to be reinterpreted in some sensible and pedagogically defensible way.

A paper by Elaine Wylie and the present writer on writing proficiency scales is contained in the portfolio of papers. The scales could either be an adaptation of existing scales or be independently developed but should, in any case for reasons of compatibility and communication, be able to be related to the levels of one or more of the established and widely used scales. The development of the scales should be undertaken by a group including native speaking teachers of Japanese as a Second or Foreign Language, native English speaking teachers of Japanese with a high level of proficiency in Japanese, and one or more experts in proficiency and its assessment. The scales should be both informally and formally trialled, presumably in much the same way as the ASLPR was trialled [see Ingram 1984 in the annotated bibliography].

At the time when preparation of this paper was about to commence, the Chairman of one of the language committees of NAATI (the National Accreditation Authority for Translators and Interpreters) suggested that there was no need for work on new proficiency assessment procedures in Indonesian, Japanese or any other language since the NAATI scales and assessment procedures were suitable for the purpose. Such a view misrepresents not only the nature of language proficiency but also the nature of interpreting and translating and suggests a lack of familiarity with the aims and methods of language testing. Indeed, the proposal that NAATI was the appropriate place for the location of some sort of authority for proficiency assessment seems to suggest a view that proficiency assessment is best carried out through translation exercises and, in any case, that translation and interpreting do not involve any skill beyond language proficiency. On the contrary, interpreting and translating are special skills which undoubtedly require proficiency but, in addition, require distinctive interpreting and translating skills, complemented in the profession by appropriate ethical and other work practices. The NAATI scales necessarily take the specific interpreting and translating skills

into account and, even if in every other way they were satisfactory as descriptions of proficiency, the inclusion of these interpreting and translating dimensions invalidate them as proficiency scales. The proposal that NAATI should become the administrative authority for language proficiency tests is unsatisfactory also since, first, if there are persons with expertise in language testing involved with NAATI, they do not publish in the area and, second, proficiency assessment is only one aspect of the array of language tests that are required and it sits more appropriately with the rest of the field of language testing than it does in the specialist field of interpreting and translating.

Industry: There are many uses for tests of Japanese proficiency in industry and commerce. In vocational situations, it is generally not an employee's formal knowledge of the language that is important but whether the applicant has sufficient proficiency to be able to carry out certain tasks (e.g., translate or gloss a letter from a Japanese agent, prepare promotional literature, talk to prospective clients in Japanese, and so on). The precise needs and the full range of needs will have to be identified by the representatives from industry present at this meeting but, in addition, as work proceeds on developing necessary tests, a proper needs analysis may be required. This undoubtedly would reinforce the need for proficiency scales but may also identify the need for other test types.

In industry, proficiency scales serve at least two sets of purposes: on the one hand, they can be used to identify the sorts of skills that employees require in order to carry out particular tasks and, on the other hand, they can be used to measure the proficiency of those employees in order to assess their suitability for employment, to assess their suitability for carrying out specified tasks when they are employed, or to assess eligibility for salary incentives or productivity payments for language skills. For these purposes, either general proficiency or special purpose scales and tests may be required and though common widely used instruments may be suitable in many cases, there will undoubtedly be instances where a specially designed scale, set of criteria or test is required. The source for such assistance should undoubtedly be one of the testing units of the National Languages Institute of Australia located in Melbourne and Griffith Universities or an accredited language assessor as discussed subsequently.

Language Programmes: Proficiency scales are not only useful in the assessment of learners' practical language skills but they have an essential contribution to make to the development, teaching and management of language programmes. This is now a matter for wide-ranging research and discussion worldwide, as some of the papers listed in the bibliographies indicate [e.g., Clark 1983] and the present writer has written at some length on the topic over the last decade [e.g., see Ingram 1986, 1985, 1984, 1979/1982]. Without going into detail, it suffices to observe that proficiency scales, because they trace the development of proficiency from zero to native-like, provide an overarching framework within which language programmes can be planned; they enable entry levels and goals to be specified precisely, identify dominant developmental trends that teaching should seek to facilitate, provide a means of assessing the suitability of activities to be proposed in a course, and enable materials to be classified and matched to the learners' stage of development. As already noted and as discussed in the present writer's report of the evaluation of the three national curriculum projects in Indonesian, Chinese and Japanese, those projects and the Australian Language Levels Project have suffered seriously from the lack of proficiency scales suitable for learners in both Primary and Secondary Schools.

Not least, the existence of a suitable proficiency rating system enables the goals to be pursued by courses to be made explicit and the success of courses in developing proficiency to be monitored. At present there is no information available on what proficiency levels students achieve in their tertiary language programmes in Australia, courses tend to meander along with exercises that may or may not be appropriate and may or may not contribute to the real development of language proficiency, and there are few if any attempts to define graduation requirements in terms of proficiency attained rather than through "time serving", i.e., doing a set number of units. Not only is this unsatisfactory in terms of not knowing what skills learners have on completion of their course, but it allows ineffective teaching and poor course design to go unchallenged.

Quality Control: One of the vital reasons to have satisfactory proficiency scales available and used in language programmes is to exert a quality control mechanism over courses and to increase accountability amongst programme planners and teachers [cf. Bachman and Savignon 1986: 380]. Accountability demands that programmes be so designed with respect to content, methods and time allocation as to enable the specified goals to be attained but if, as just noted, goals do not state unambiguously the proficiency levels to be sought and, if student assessment does not indicate the levels actually being attained, ineffective courses readily go unchecked. In addition, the end-users of the skills developed by courses (i.e., the future employers) can legitimately complain that new graduates lack the skills needed to be of use to them and this in turn puts the whole notion of language skills being vocationally relevant into question.

If quality is to be assured, then appropriate levels of language skills and professional training need to be insisted upon amongst language teachers. For this purpose, appropriate proficiency scales are needed and the minimum skills to be expected need to be specified and written into such controls as teacher registration provisions. In the recommendations for a new State language education policy in Queensland recently prepared by the Centre for Applied Linguistics and Languages at Griffith University, Glyn John and I have recommended that the minimum proficiency for teachers be set at S:3, L:3, W:3, R:3 except that, in character-based languages, it should be not less than S:3, L:3, W:2, R:2.

The fundamental reason for the need for proficiency assessment instruments to ensure quality control is that, without them, it is impossible to obtain and maintain the data on student attainment, course goals, and so on, that are needed if quality is to be assessed: without such scales, essentially, there is no means by which to talk about student performance and no means by which learners themselves can identify their own goals and work towards them systematically. Thus the existence of suitable readily comprehensible statements of proficiency that can be used to specify goals and to talk about proficiency can have a favourable effect on one's ability to monitor the effectiveness of language learning programmes, it can facilitate demands for teacher quality, and it can have a beneficial effect on learners themselves as they become more conscious of what they can do in the language, of what it means to attain proficiency in the language, of the incremental progress they are making, and of the practical usefulness of the sorts of assessment procedures they undergo in order to have their proficiency specified [see also Moeller 1988, Clark 1983, Babcock 1985, Hirsch 1985].

Diversifying and Increasing Language Learning: If the diverse needs in Australia for language skills are to be met and if individual needs and interests are to be catered for, it is essential that the whole structure of language learning be revised. Not least, we have to accept that language learning goes on in other places beside school or university classrooms. The education system needs also to start to value language skills that may have been acquired other than from formal teaching, e.g., through growing up in a bilingual home, through overseas travel, and so on. If this valuing of other language learning is to be achieved, then at least three things are needed: first, that language courses start to value proficiency attainment; second, that the means be available to assess proficiency whether or not learners have followed a formal course in schools, universities or elsewhere; and, third, that suitable proficiency scales be available and their use built into the regular institutional assessment requirements. In addition, if things other than proficiency are to be valued and built into assessment requirements, then tests for these also need to be devised so that outside learning can be assessed and recognized by the student's institution. Again, the recommendations for a new State language education policy in Queensland include proposals to enable this recognition of outside learning to occur and, thereby, to enable language learning opportunities in all languages to be increased and diversified. Again, however, the key to this is the availability of suitable proficiency rating instruments.

Attainment Tests: The distinction was made earlier between proficiency and course attainment (see also "Assessing Proficiency", Section IV.7). It is an important distinction which is often overlooked by teachers who assume that what they teach in class is automatically going to be available to enhance their students' proficiency or that each item of vocabulary, structure or function learned is going to make a significant difference to the learner's proficiency. It is for this reason that one not infrequently hears the complaint that proficiency instruments are too blunt: this criticism often arises from a failure to distinguish between course attainment or the internalization of what is formally taught and learned in class, from proficiency which exists independently of the courses followed. In that attainment tests are based on the content of the actual courses taught, their construction would seem less relevant to the concerns of the Working Party on the Assessment of Proficiency. However, where students are following courses based on the national curriculum, it would seem advantageous for professionally designed attainment tests to be available both because this would ensure that they were more competently designed and because it would relieve teachers of the necessity of designing their own tests [cf. Clark 1983: 441].

Streaming: Similar needs and answers exist in the area of streaming as for Attainment Testing. Though proficiency may be a factor in streaming, in "Assessing Proficiency" it is argued that course related tests for streaming purposes are probably more appropriate. Where the courses available and into which students are to be streamed are based on a common curriculum such as the national curriculum, there would be advantage in having professionally designed tests available. Thus, for instance, the Stages that are specified in the ALL Project material and in the national curricula in Asian languages (including Japanese) might be accompanied by stage-based tests that can be used for streaming purposes (though, as has been noted in reviews of the ALL Project and of the national curricula, the specification of the Stages is such that it is difficult to know what proficiency levels students entering the Stage will have.

Diagnosis: Diagnostic testing aims to identify the strengths and weaknesses in a learner's language so that weaknesses identified can become the focus of specific intervention by the teacher or learner. To some extent, diagnostic tests are related to the actual content of a course but insofar as the students are following courses based on the national curriculum and insofar as one accepts that the broad schedule of development in a language is more or less universal, there would seem some advantage in having available a battery of diagnostic tests that are professionally constructed and assist teachers in the course of their regular classwork to diagnose learners' strengths and weaknesses as a basis for remedial intervention. Such tests would need to cover all those aspects of the language and its use that occur in language programmes including the language elements of syntax, vocabulary, cohesion, etc and functional features [cf. Clark 1983].

Self-Assessment: One of the benefits that has flowed from the development and use of proficiency scales is that of motivation resulting from students' being more aware of learning goals and of increases in their proficiency and from their increased ability to monitor their own progress. Self-monitoring is particularly important in the context of diversifying language learning opportunities especially in order to include self-access and autonomous learning. For this purpose, it is desirable to develop tests, scales and procedures that can be used for self-assessment. Proficiency scales for use in self-assessment would be based on the scales for teacher use but would be simplified and be accompanied by additional practical guidance (e.g., questionnaires and examples) to assist learners in introspecting their language behaviour. The scales and other tests would desirably be related eventually to national curricula, to distance education courses, and to self-access materials so as to guide learners through available learning materials.

Self-assessment procedures involving both assisted self-rating scales and other test types could also be incorporated into computer software and interactive video for use in self-access centres, in learning at home, and for assessment purposes in industry. Such procedures could involve all four macroskills, both general and special purpose proficiency, and be interfaced with self-access courses, remedial materials, and distance learning programmes so that the learners can both self-assess and self-direct through the available learning materials [cf. Davies 1984: 48].

Classifying Materials and Activities: One of the advantages of having widely accepted and readily used proficiency scales available is that they facilitate the classification of library materials and other teaching materials. Because proficiency scales describe language behaviour, they allow one to classify materials according to the language behaviour (i.e., the proficiency level) required to cope with them. The ASLPR is already used for this purpose both in language centres and in other libraries and proficiency scales specifically for Japanese would similarly assist. In addition, since proficiency scales allow one to specify entry levels to courses and their goals and to assess the level of language performance required to cope with materials and activities, they provide a way of assessing the suitability of materials and activities to learners and their stage of development. Thus they have an important role in facilitating good language teaching sensitive to the needs and capacities of the learners. Too often when proficiency scales are not available or are not used, teachers adopt materials and activities inappropriate to the level of their students. The most frequent example of this has been the premature introduction of literary studies into language courses at

elementary levels. Other examples exist in the national curricula recently reviewed, where, in some languages (Japanese was less of a problem in this regard), reading and writing activities involving characters were proposed that were not at all compatible with the likely proficiency levels of the learners; likewise, the ALL Project materials in general and the curricula based on them are notably deficient in relating activities to the level of the learners as a result, in particular, of their failure to acknowledge the pivotal role of proficiency in language programmes or to utilize proficiency scales.

Programme Rationalisation: At present, little information is available on how rapidly learners of different ages progress through different proficiency levels in Japanese and other languages. The only tables that relate proficiency levels to course lengths are those published by the United States Foreign Service Institute School of Language Studies and a comparable table from Australian sources now publicly available in the recently released *French in Australia* written by Joe Lo Bianco and Alain Monteil. However, these figures relate to languages learned in circumstances that make it difficult confidently to generalize the expectations to courses in the formal education system. Yet, without such information, programme planning is largely based on guesswork, in general administrators and other budget managers grossly over-estimate what can be achieved, and there is urgent need for more accurate and generalizable data to be accumulated. For it to be accumulated, access is required to nationally accepted proficiency scales and the widespread maintenance of records of entry and exit levels, course lengths and types, and student characteristics. The more rational planning of educational provisions and of language learning programmes in particular that would result from the availability of such data would itself more than justify the investment of time and effort required to develop suitable proficiency scales and to maintain the necessary records. Undoubtedly one of the reasons why this sort of data has not been accumulated in the past has been the fear of teachers that it would be used to make them more accountable for the quality of their programmes and the effectiveness of their teaching and, indeed, when, as academic advisor to the Adult Migrant Education Program some ten years ago, I advocated that records be maintained for using the ASLPR, the suggestion was greeted with hostility. For educational planners and language policy-makers, however, the availability of such data would be of very considerable value and, in fact, because most policy-makers and decision-makers outside of the language teaching profession seem to over-estimate the rate at which proficiency changes and hence under-estimate the resources that should be allocated, I believe that language teachers (especially teachers of languages such as Japanese) have more to gain than to lose from ensuring that this sort of data is accumulated. There would, I believe, be much advantage for all concerned (learners, teachers, educational institutions, the Asian Studies Council and others) if one of the outcomes from this Working Group on Assessment of Proficiency Levels for Students of Japanese was not only that suitable proficiency measurement instruments would be developed but that a viable approach to accumulating the sort of data discussed here should be developed.

Other Tests: Reference has been made to the need for proficiency tests in Japanese, attainment tests and diagnostic tests. However, it would be desirable also to consider whether other types of tests might not also be required. One of the most obvious of these is aptitude tests.

The whole field of language aptitude testing has been relatively neglected since the early 1960's when Carroll developed the MLAT which remains the principal test available. Consequently, the aptitude tests available tend not to be satisfactory in the context of modern approaches to language teaching and learning and are excessively constricted in their concept of what aptitude involves. Probably the outcome of a systematic attempt to develop alternative aptitude instruments would not be a single aptitude test analogous to the MLAT but a profiling battery that would take into account the traditional factors such as aural acuity, grammatical sensitivity, short term and longer term memory, but also factors such as social and psychological distance, introversion-extroversion, age-related factors, motivation, attitudes, previous language learning experiences, and so on. The need for such tests is evident especially in circumstances where employers in industry or administrators in education are allocating resources to enable employees to re-train: it is appropriate, before people are selected for such programmes, that their likely ability to succeed in the time available be assessed so that the available places can be directed towards someone else or the available time and resources adjusted to enable them to succeed.

Other tests that warrant consideration include tests of cross-cultural attitudes and of cultural understanding though this meeting may also identify others.

Research: There has been a vast amount of research into language learning and teaching over many years (indeed for several thousand) years and this has proliferated over the last three decades as applied linguistics assumed its role as a distinctive field. However, most of that research has been based on the teaching of English and to a lesser extent on a few European languages while the distinctive problems of the teaching and learning of Asian languages and of Japanese in particular have received comparatively little attention.

The development of valid and reliable proficiency assessment procedures is a pre-requisite for much research into the teaching of Japanese since the purpose of such research is to enhance the efficiency and effectiveness of the teaching and this implies the possibility of being able to make accurate assessments of proficiency changes achieved by alternative methods.

In addition, the problem of developing suitable test instruments (whether direct, indirect or semi-direct) will raise issues that are distinctive for Japanese at least in comparison with European languages. Not least amongst these are the implications of a very different culture and the different linguistic, paralinguistic and other sociocultural ways in which human relationships and interactions are mediated.

Of particular significance for research is the need to better understand the problems inherent in the teaching and learning of character-based writing systems, how reading and writing can most effectively and efficiently be taught, whether there is, in fact, any value in commencing their teaching using Romanized alphabets as initial teaching alphabets, and whether the teaching of reading and writing can be facilitated by the use of such word-processing packages as *Jieijing*, recently developed by Dr Eric Chappell of Griffith University and already being adopted in many countries. In addition, of course, developments in our understanding of the teaching of reading and writing in character-based languages will have important implications for their testing leading to a re-assessment of the

whole range of test types for reading and writing and undoubtedly to the development of new types.

Training in Testing: Of no lesser importance than the development of suitable instruments for testing in Japanese is the need for much better training in language testing amongst teachers of Japanese and, for that matter, all other languages. I started the complete version of the paper "Assessing Proficiency" which is in the portfolio of papers for this conference, by saying that no activity in language teaching was more widely practised and practised so badly as language testing. Regrettably, the situation is little changed some seven years later not least because, in Australia at least, language testing is generally accorded a few short hours in a teacher education course, there have been no specialist courses in language testing available, few if any graduate students have conducted research in language testing in Australia, and, if the literature searches I conducted were adequate, no one has done so in any aspect of the testing of Japanese. This is not just the fault of language teachers or the training programmes, however, since, to be frank, language testers with their traditional obsession with psychometrics rather than psycholinguistics, methodology and syllabus design, have too often seemed irrelevant to the real processes of language teaching and learning. This is, however, no longer the case, especially since recent interest in the concept of proficiency and its assessment emerged in the 1970's. It is essential that, in future, pre-service and in-service teacher education courses for language teachers give due attention to language testing and perhaps one of the outcomes of this meeting could be recommendations to this effect.

Since the early '80's, I have pointed out in numerous papers (including in "Assessing Proficiency") the need for what I have been calling "accredited language assessors". Even if all language teachers were to receive thorough training in language testing, there would, I believe, still be a need for accredited language assessors, not least in Japanese. The fact of the matter is that everyday scores of decisions are made about the proficiency levels of individuals in many different languages and about the language skills that are needed to carry out particular tasks. At present, these decisions are most often made by untrained laypersons with little formal knowledge either of the nature of language or of language assessment. Yet these decisions are often of very great importance ranging from whether or not a person may be accepted into a language programme to legal responsibility to selection for employment, and so on. There is need for individual speakers of a language, for employers, educational institutions, and courts of law amongst many others to have access to persons who are trained both to assess a person's proficiency in a language and to advise on the proficiency required for various purposes. Just as there is established in many countries including Australia an accreditation system for translators and interpreters, I believe there is such a need for language assessors and for similar reasons: the decisions that are made concerning a person's proficiency are often of far-reaching importance and all persons should have the right to valid and reliable assessment and to an appeal process by which injustices if they occur can be rectified. Such language assessors would not necessarily be language teachers but would be graduates with some training in the linguistic sciences, would have taken at least a three month course in all aspects of language testing, and would have demonstrated their skill in identifying language needs and in all forms of language testing (in particular proficiency assessment). A considerable component of such training would, of course, be specific to the language in question, in this case to Japanese.

V TEST TYPES: The types of test instruments that are required to meet the needs of testing in Japanese, especially of proficiency measurement in Japanese, are already indicated in the outline of needs just provided. In brief, they include amongst others:

- * proficiency scales (general, specific, global and analytic)
- * "free" and controlled interview and other procedures for the application of proficiency rating scales and the scoring and interpretation of test scores
- * test batteries (perhaps analogous to IELTS in English) that enable large scale proficiency assessment in circumstances where one can have less confidence in the training and linguistic and professional competence of the test administrators
- * alternatives forms of Speaking tests, e.g., Simulated Oral Proficiency Interviews [cf. Stansfield 1989, Stansfield and Kenyon 1988, Clark 1986]
- * training procedures for interviewers and assessors using proficiency scales
- * self-assessment instruments and procedures
- * attainment tests
- * diagnostic tests
- * possibly other standardized tests aimed at measuring proficiency or some aspect of it
- * item banks (validated by trialling and item response theory and related to proficiency levels) that can be drawn on by teachers, for the construction of more widely used tests, and for inclusion in controlled interview procedures
- * tests of specific features of the language, e.g., vocabulary, grammar, functions, cohesion, etc.
- * computer based tests

VI TESTS OF JAPANESE: In the bibliographies, I have included the names of whatever significant tests and proficiency scales of Japanese that I was able to find. In addition to these, more general scales such as the FSI Scale, ILR Scale and the ASLPR are available. In 1982, the ASLPR was also provided in versions for French, Italian and Japanese mainly as an exercise in demonstrating that it was equally applicable to languages other than English. However, Elaine Wylie and I were not very confident of either the examples provided by our Japanese informant or that the issues distinctive to Japanese had been adequately handled and we have been reluctant to release the scale for use despite reasonably successful trials. It was included in the portfolio solely to exemplify the possibility of providing a language-specific version of a general scale. Since the listing in the bibliographies also provides a short description of the scales, no further assessment will be provided here except to say that there is clearly great

need for extensive test development activity in the area of testing Japanese as a foreign language.

VII THE FUTURE: The principal purpose of this conference is to consider the needs for tests of Japanese, especially in the area of Japanese proficiency assessment though it is not possible, in doing so, to exclude consideration of needs for other tests. In the "Advance Organizer" sheet distributed with the portfolio of papers, I raised a number of questions that might direct our thoughts into the area of needs and possible action and in this paper I have sought to canvass what I, as an applied linguist with interests especially in testing, policy-making and curriculum development, see as areas of needs for Japanese testing. The rest of today is intended to consider developments in this area elsewhere and to hear from representatives from industry (the end-users of the products of our courses and of our test results) what they see as the needs for Japanese proficiency testing; over the next two days, those testing experts who have been invited to this conference are charged with specifying needs, identifying what tests are needed to respond to those needs, and outlining in some detail, costing, and prioritizing possible test development projects. Consideration of the projects should include detail of how the tests are to be developed, what documentation is needed, and what validation procedures are required. The comprehensive development process of the IELTS as outlined in one of the papers in the portfolio may be suggestive in this regard and it is worth noting the stages that were undertaken in which needs were identified, the purpose of the test and sub-tests specified, the types of candidates identified, item types considered, test specifications written, item writing teams established, training programmes for test administrators prepared, and comprehensive trialling for purposes of standardization undertaken.

As test development projects are identified over the next two days, it will be necessary to consider how the projects might be undertaken, what role the National Languages Institute of Australia and, in particular, its test centres might have and what input is possible and desirable from other countries, especially those represented here. It will also be necessary to consider how, in the longer term, the tests developed are to be distributed, managed and administered. Some tests, for example, might be sold or otherwise made available to teachers, institutions or accredited language assessors; other tests might be centrally administered, for example by the NLIA; item banks might be stored, perhaps at one of the test centres of the NLIA, and we might consider the need for the Clearinghouse of the NLIA at the University of Queensland to house a databank of tests and items, a possibility that was recently discussed with Roly Sussex, the Director of that Centre, as an activity that the Testing and Curriculum unit at Griffith University might undertake.

Undoubtedly the activity foreshadowed by this meeting will constitute a major contribution to research in the area of the teaching of Japanese as a Foreign Language and, in particular, in the area of Japanese language testing. It will complement the not inconsiderable research and development activity that has gone on in other areas of testing, not least in that of proficiency in English as a Second or Foreign Language, and it provides an opportunity to relate some of that work to the testing of Japanese and to confront some of the distinctive issues affecting Japanese itself, other Asian languages, and, not least, character-based languages. Granted the economic and political importance of Japan to Australia and the increasingly close ties between the two countries, it is appropriate that

such research should assume a significant place in Australian applied linguistic activity.

-000-

-000-

-000-

INTERNATIONAL WORKING GROUP ON ASSESSMENT OF PROFICIENCY LEVELS FOR
STUDENTS OF JAPANESE AS A FOREIGN LANGUAGE

CONTENTS OF PORTFOLIO

Advance Organizer

Bibliographies

- Clark, John L. D. 1986. *Handbook for the Development of Tape-Mediated, ACTFL/ILR Scale-based Tests of Speaking Proficiency in the Less Commonly Taught Languages*. Washington, D.C.: Center for Applied Linguistics.
- Ingram, D. E. 1990. "The International English Language Testing System (IELTS): Its Nature and Development". Paper to the 1990 RELC Seminar, Regional Language Centre, Singapore, 9-12 April, 1990. (mimeograph)
- Ingram, D. E. 1990a. "The International English Language Testing System (IELTS): The Speaking Test". Paper to the 1990 RELC Seminar, Regional Language Centre, Singapore, 9-12 April, 1990. (mimeograph)
- Ingram, D. E. 1985. "Assessing Proficiency: An Overview on Some Aspects of Testing". In Hyltenstam and Pienemann 1985: 215-276.
- Ingram, D. E. and Elaine Wylie. 1989. "Developing Proficiency Scales for Communicative Assessment". An abridged version of a paper presented at the National Assessment Consultation for the National Assessment Framework for Languages at Senior Secondary Level, Sydney, 5 December, 1989.
- Ingram, D. E. and Elaine Wylie. 1979/1985. *The Australian Second Language Proficiency Ratings*. (mimeograph)
- Ingram, D. E. and Elaine Wylie. 1982. *The Australian Second Language Proficiency Ratings (Version for Japanese as a Foreign Language)*. (mimeograph) [NB. Early draft only - not released for use or citing].
- Morrow, Keith. 1990. "Evaluating Communicative Tests". Paper to the 1990 RELC Seminar, Regional Language Centre, Singapore, 9-12 April, 1990. (mimeograph)
- Stansfield, Charles W. 1989. *Simulated Oral Proficiency Interviews. ERIC Digest*, December 1989. Washington, D.C.: Center for Applied Linguistics.
- Thompson, Irene, Richard T. Thompson and David Hiple. 1987. "Adapting the ACTFL/ETS Guidelines to the Less Commonly Taught Languages". In Stansfield and Harman 1987, Chapter 4.
- Thompson, Richard T. and Dora E. Johnson. 1988. *ERIC Digest: Proficiency Testing in the Less Commonly Taught Languages*. Washington, D.C.: Center for Applied Linguistics.

INTERNATIONAL WORKING GROUP ON ASSESSMENT OF PROFICIENCY
LEVELS FOR STUDENTS OF JAPANESE AS A FOREIGN LANGUAGE

BIBLIOGRAPHIES

compiled by

D. E. Ingram

Professor of Applied Linguistics and Director,
Centre for Applied Linguistics and Languages,
Griffith University, Brisbane, Australia.

NOTE: Searches were conducted both through Australian libraries and through the Center for Applied Linguistics in Washington, D.C.. Very few items were found linking Japanese and language testing. The following very selective bibliographies are not intended to be comprehensive (the literature is vast) but they indicate some items in the general area of language proficiency and its assessment that may be useful as work is commenced on particular projects. Additional more specific searches or a re-reading of the ones made for this conference should be carried out as each particular project is commenced. The bibliographies provided are in four sections: A. Ready-made bibliographies from various sources; B. an Annotated Bibliography; C. a General Bibliography; D. Relevant Tests, Scales, etc. A portfolio of papers presenting basic issues is also provided though copyright and other considerations has necessarily limited the range from which they could be chosen.

A. READY-MADE BIBLIOGRAPHIES

1. ERIC Clearinghouse on Languages and Linguistics. 1990. Computer Search No. 917: *Oral Proficiency Testing*. Washington, D.C.: Center for Applied Linguistics.
2. Regional Language Centre. 1990. *Specialized Bibliography 48: Language Testing and Language Programme Evaluation*. Singapore: RELC.
3. ELTS Revision Project Team. 1989. "Bibliodat", bibliography compiled in the course of the ELTS Revision Project and development of the International English Language Testing System. Cambridge: UCLES. (mimeograph)
4. Galloway, Vicki, Charles W. Stansfield, and Lynn E. Thompson. 1987. "Topical Bibliography of Proficiency-Related Issues". In Stansfield and Harman 1987, Chapter 6.

B. ANNOTATED BIBLIOGRAPHY

Alderson, J. Charles, Karl J. Krahne and Charles W. Stansfield. 1987. *Reviews of English Language Proficiency Tests*. Washington, D.C.: TESOL.

This volume contains reviews of some 46 tests of English used for a variety of purposes including proficiency assessment. Though the views expressed are those of the reviewers (which, in the opinion of the volume's editors, sometimes significantly qualifies the worth of some reviews and their insightfulness), it is a useful reference on the range of test types available and the strengths and weaknesses to be found in tests.

Allen, J.P.B. and Alan Davies. (eds.). 1977. *Testing and Experimental Methods. The Edinburgh Course in Applied Linguistics. Vol. 4*. London: O.U.P.

This is a standard reference in language testing though with little on the concept of proficiency and its assessment. It has chapters on basic concepts in testing, the construction of language tests, the design and interpretation of experiments, and statistical procedures.

Bachman, Lyle F. and Adrian S. Palmer. 1981 . "The Construct Validation of the FSI Oral Proficiency Interview". In *Language Learning*, Vol. 31, No. 1, 1981, pp. 67-86.

The authors examine the construct validity of the FSI oral interview using a multitrait-multimethod matrix comprising six measures. The results indicate both convergent and discriminant validity for the FSI interview and the authors conclude that the results provide strong evidence against a model of unitary language competence and for a partially divisible language competence.

Barnwell, David. 1986. "Who is to Judge how well Others Speak? An Experiment with the ACTFL/ETS Oral Proficiency Scale", paper to the Eastern States Conference on Linguistics. ED296589 FL017471.

Seven teachers of Spanish were given brief formal training on the use of the ACTFL Scale and then they evaluated recorded interviews of learners of Spanish at different proficiency levels. A high degree of agreement amongst the raters was found despite their minimal training.

Biersteker, Ann. 1986. "A Language Proficiency Testing Model for Business Needs". In *Proceedings of the Fifth Annual Eastern Michigan University Conference on Languages for Business and the Professions*. ED295447 FL016621.

The ACTFL OPI was adapted to focus on business contexts with proficiency profiles adapted accordingly. These were found to be more specific in identifying skills and needs for business purposes and suggested proficiency guidelines are proposed.

Byrnes, Heidi and Michael Canale. 1987. *Defining and Developing Proficiency: Guidelines, Implementations, and Concepts*. Lincolnwood, Ill.: National Textbook Company.

Three of the five chapters in this volume concern the ACTFL Guidelines while the fourth concerns the application of proficiency levels to classifying texts and the last discusses second language acquisition and the implications of proficiency for course design and teaching methods.

Canale, Michael and Merrill Swain. 1980. "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing". In *Applied Linguistics*, Vol. 1, No. 1, 1980, pp. 1-47.

This paper examines the concept of communicative competence looking at some of the theories that have been produced. The authors propose a theoretical framework for communicative competence and examine its implications for foreign language teaching and testing.

Carroll, Brendan J. 1980. *Testing Communicative Performance*. Oxford: Pergamon.

Carroll discusses the assessment of communicative performance and the design, development, and uses of communicative tests. The last chapter reviews some relevant literature in the area of communicative testing. Lengthy appendices include a needs profile, "information banks" for test development (actually specifications for an RSA test), and sample rating scales.

Clark, John L.D. 1988. "Validation of a Tape-Mediated ACTFL/ILR-Scale Based Test of Chinese Speaking Proficiency". In *Language Testing*, Vol. 5, No. 2, December 1988, pp. 187-205.

This paper reports a study in which students' performance on a simulated oral proficiency test is compared with their performance on a live interview. Results were similar though candidates found the simulated interview more difficult.

Clark, John L. D. 1986. *Handbook for the Development of Tape-Mediated, ACTFL/ILR Scale-based Tests of Speaking Proficiency in the Less Commonly Taught Languages*. Washington, D.C.: Center for Applied Linguistics.

This small booklet documents the development of a tape-based test for Chinese. The test is a "proficiency-oriented semi-direct test" which approximates to the interview situation but the stimuli and the candidate's responses are recorded and the candidate's performance is rated using the ACTFL/ILR Proficiency Guidelines. The procedures and issues raised are relevant to other languages and may be especially useful where live interviewing is not possible.

Clark, John L. D. (ed.). 1978. *Direct Testing of Speaking Proficiency: Theory and Application*. Princeton, NJ: Educational Testing Service.

This volume contains eighteen papers from a two-day conference on the assessment of speaking proficiency. Most interest focussed around the development and use of the FSI Scale and the associated interview procedures. Other topics included the use of the FSI Scale and interview in different contexts for different purposes including diagnostic evaluation, teacher certification, and High School attainment.

Clark, John L. D. 1972. *Foreign-Language Testing: Theory and Practice*. Philadelphia, Pa.: Center for Curriculum Development.

This volume considers different types of testing including prognostic testing and tests of achievement (all four macroskills), knowledge (including culture and literature), and proficiency (with special reference to the FSI Scale and interview). It also discusses the range and roles of published tests.

Clark, John L. D. and Ray T. Clifford. 1988. "The FSI, ILR, ACTFL Proficiency Scales and Testing Techniques: Development, Current Status, and Needed Research". In *Studies in Second Language Acquisition*, Vol. 10, No. 2, June 1988, pp. 129-147.

Describes the rationale, history, and current state of the proficiency movement in the U.S. and identifies areas requiring further research and development.

Clark, John L. D. and Eleanor H. Jorden. 1984. *A Study of Language Attrition in Former U.S. Students of Japanese and Implications for Design of Curriculum and Teaching Materials: Final Project Report*. Washington, D.C.: Center for Applied Linguistics.

This publication reports a long-term study on the attrition of skills in Japanese amongst university students of Japanese. The period of attrition differed. Initial proficiency (measured using the Japanese Proficiency Test and a self-assessment schedule) seemed to have little influence on the extent of attrition, decreases in proficiency were inversely related to the amount of use made of the language, attrition seemed to be skill-specific, and re-acquisition of skills seemed to be faster using familiar rather than unfamiliar materials. A number of case studies are provided and confirmed the group data analyses.

Cziko, Gary A. 1981. "Psychometric and Edumetric Approaches to Language Testing: Implications and Applications". In *Applied Linguistics*, Vol. 2, No. 1, Spring 1981, pp. 27-44.

The essential differences between psychometric and edumetric tests in education are outlined, implications drawn for language testing, discrete-point and integrative approaches reviewed, and the construction and use of edumetric integrative tests are described.

Davies, Alan. 1984. "Computer-Assisted Language Testing". In *Calico Journal*, Vol. 1, No. 5, June 1984, pp. 41-42 and 48.

This short paper reviews the state of computer-assisted language testing in the first part of the 1980's and identifies its strengths and limitations. It identifies three areas of particular interest: testing for self-access purposes, cloze and reading-speed tests, and diagnostic testing.

Doyle, Michael S. 1988. "Business Spanish: Role Playing for Oral Proficiency". In *Proceedings of the Seventh Annual Eastern Michigan University Conference on Language for Business and the Professions*. ED304938 FLO17895.

This reports the use of ACTFL OPI situation cards in special purpose areas of the language for both measurement and teaching purposes.

Dugan, J. Sanford. 1988. "Standardized Tests as an Alternative to the Oral Interview". In *Proceedings of the Seventh Annual Eastern Michigan University Conference on Languages for Business and the Professions*. ED304923 FLO17880.

A moderate correlation was found between the MLA listening and speaking tests and the ACTFL/ETS OPI results. Problems in interpreting the MLA results were also noted.

Ellis, Rod. 1984. "Communication Strategies and the Evaluation of Communicative Performance". In *ELT Journal*, Vol. 38, No. 1, January 1984, pp. 39-44.

Ellis proposes that, instead of trying to define proficiency or communicative competence, the tester should focus on "communication strategies" for assessment purposes, i.e., strategies adopted by learners in order to communicate when they lack the necessary language.

Frith, James R. (ed.). 1980. *Measuring Spoken Language Proficiency*. Washington, D.C.: Georgetown University Press.

This small volume contains papers relevant to the measurement of speaking proficiency. It considers factors in speaking proficiency, ways of measuring speaking proficiency using both live and recorded interviews, and examples of the use of oral proficiency ratings.

Green, Peter S. (ed.). 1987. *Communicative Language Testing: A Resource Book for Teacher Trainers*. Strasbourg: Council for Cultural Cooperation.

This book is intended as a resource for teacher trainers who wish to organize inservice programmes on the communicative testing of foreign language competence in High School students. The proposals are organized under six themes including the nature of communication, the content of a communicative testing syllabus, testing communicative competence at a national level, classroom assessment, teacher training in testing communicative competence, and testing communicative performance.

Greenberg, Karen. 1986. "The Development and Validation of the TOEFL Writing Test: A Discussion of TOEFL Research Reports 15 and 19". In *TOEFL Quarterly*, Vol. 20, No. 3, 1986, pp. 531-544.

This paper reviews the research conducted prior to the introduction of a writing test in TOEFL, looking especially at TOEFL Research Report No. 15 (*Survey of Academic Writing Tasks Required of Graduate and Undergraduate Foreign Students*) and TOEFL Research Report No. 19 (*Relationship of Admission Test Scores to Writing Performance of Native and Nonnative Speakers of English*). The paper outlines the writing tasks identified and questions the choice of tasks on which TOEFL focusses. At the same time the author supports the use of direct assessment approaches.

Griffin, Patrick et al. 1988. "An Algorithmic Approach to Prescriptive Assessment in English as a Second Language". In *Language Testing*, Vol. 5, No. 1, June 1988, pp. 1-18.

Describes the development using specially developed algorithms of interview items for use with the ASLPR in the range of 0 to 1+.

Harris, David P. 1969. *Testing English as a Second Language*. New York: McGraw-Hill.

This book pre-dates developments in the concept and testing of proficiency and is included, like the books by Valette and Lado, as a standard reference in more traditional aspects of language testing. It discusses basic concepts such as validity and reliability and testing the elements of the language (grammar, pronunciation, vocabulary, etc) as well as the four macroskills. It has chapters on constructing tests, administering tests, interpreting and using test results, and basic statistics.

Heaton, J. B. 1975/1988. *Writing English Language Tests*. London: Longman.

A general and fairly traditional text on language testing with most emphasis on objective tests. It contains a general discussion of language testing and focusses on testing the elements of the language and the macroskills in six of the ten chapters. Other chapters consider standard test evaluation concepts and procedures and the interpretation of test scores.

Henning, Grant. 1983. "Oral Proficiency Testing: Comparative Validities of Interview, Imitation, and Completion Methods". In *Language Learning*, Vol. 33, No. 3, 1983, pp. 315-332.

Henning compares three oral testing methods: imitation, completion and FSI-type interview. Rankings in terms of overall validity across all indexes used were pronunciation, interview and completion. Regression analysis showed that the interview was more closely related to the grammar skill.

Higgs, Theodore V. (ed.). 1985. *Teaching for Proficiency, the Organizing Principle*. Lincolnwood, Ill.: National Textbook Company.

The main focus of the papers in this volume are on the application of the concept of proficiency and its assessment to language teaching though there are also papers on testing. The background and origins of the ACTFL Proficiency Guidelines are traced and the appendix reproduces the Guidelines and the ILR level definitions for speaking.

Ingram, D. E.. forthcoming. "Assessing Special Purpose Proficiency", Chapter One of Ingram, D. E.. in preparation. *Direct Proficiency Assessment*.

This paper discusses the nature of special purpose proficiency and its development and considers the relevance to it of direct proficiency assessment using rating scales. It proposes the development of a special purpose version of the Australian Second Language Proficiency Ratings in which the "General Description" column is the same but there is an additional column describing features specific to the special purpose area in question, an Example column and a Comment column.

Ingram, D. E. 1986. "Proficiency in Syllabus Design and Assessment". Paper to the National Conference and Seminar of the New Zealand Association of Language Teachers, Dunedin, 13-15 May, 1986. ED276246 FL015792

This paper discusses the relevance of the concept of proficiency to syllabus design and assessment giving emphasis to identifying the learner's real needs and the precise specification of goals in proficiency terms.

Ingram, D. E. 1984. *Report on the Formal Trialling of the Australian Second Language Proficiency Ratings (ASLPR)*. Canberra: DIEA/AGPS.

This publication reports the results of the formal trials of the ASLPR which examined the reliability of ratings assigned by native and non-native English speaking raters and the relationship between scores on the ASLPR and in pencil-and-paper tests.

James, Charles J. (ed). 1985. *Foreign Language Proficiency in the Classroom and Beyond*. Hastings-on-Hudson, N.Y.: ACTFL.

Proficiency and its meaning for teachers, curriculum developers, researchers, publishers, and business people are discussed dealing with such issues as definition of proficiency, curriculum and materials development, assessment, and applications beyond the classroom.

James, Gregory. 1984. "Development of an Oral Proficiency Component in a Test in English for Academic Purposes". Based on a paper to the 1984 AILA Congress, Brussels. ED256153 FLO14895.

A test of oral proficiency in academic contexts is described. The test, part of the TEEP Test, uses a prerecorded test to which the candidate listens and responds preferably in a language laboratory. Candidates demonstrate their proficiency in free speech, oral summary of material presented graphically, asking and answering questions, responding to remarks and situations, and responding to and expressing opinions.

Jones, Randall L. and Spolsky, Bernard. (eds.). 1975. *Testing Language Proficiency*. Washington, D.C.: Center for Applied Linguistics.

These are the proceedings of the 1974 Washington Language Testing Symposium which focussed especially on language proficiency assessment as it relates to the use of foreign languages at work. Its sections include papers on instances where proficiency testing is used and on assessment of proficiency in speaking, listening and reading..

Lado, Robert. 1961. *Language Testing*. London: Longmans.

This book pre-dates developments in the concept and testing of proficiency and is included, like the books by Harris and Valette, as a standard reference in more traditional aspects of language testing. It was published at the height of audiolingualism, behaviourism and structural linguistics and is a standard reference for test types of the era. Its first part discusses a theory of foreign language testing and the basic criteria for the evaluation of tests. The second part has chapters focussing in great detail on the testing of elements of the language (e.g., pronunciation, grammar, vocabulary) and the third part focusses on the macroskills. Part Four considers the testing of cross-cultural understanding and attitudes, while Part Five discusses in some detail such concepts as norms, validity, reliability, item analysis and scoring, has a chapter on different purposes for testing (e.g., diagnostic, achievement and aptitude tests), and one on designing experiments in foreign language learning.

Liskin-Gasparro, Judith E. 1984. "Foreign Language Oral Proficiency Testing: An Organizing Principle for Instruction". Paper to the Annual Meeting of the National Council on Measurement in Education, New Orleans. ED263787 FL015412.

This paper refers to programmes in which proficiency scales have been used to modify the nature of the syllabus and how it is taught using the ACTFL/ETS guidelines as an organizing principle.

Magnan, Sally S. 1988. "Grammar and the ACTFL Oral Proficiency Interview: Discussion and Data". In *Modern Language Journal*, Vol. 72, No. 3, Fall 1988, pp. 266-276.

Discusses the role of grammar in the ACTFL Scale and provides data on the grammatical accuracy of 40 interviews at various proficiency levels of French.

Moeller, Paulette. 1988. "Quality Control and French Proficiency". In *Education Canada*, Vol. 28, No. 2, Summer 1988, pp. 4-9.

This paper proposes the use of the ACTFL Scale and interview as quality control measures for students and teachers in Canada.

Munson, Johanna and Mary Harrison. 1990. "Review of the Japanese Language Proficiency Test", MATFL language testing seminar paper, Monterey Institute of International Studies. (mimeograph)

This short paper reviews the Japanese Language Proficiency Test [see Section D. Relevant Tests, Scales, etc. below], highlights its limitations as a proficiency test, and points to some construction faults in the items in the version reviewed. It draws attention to the test's failure to measure productive skills.

Oller, John W. and Kyle Perkins (eds.). 1980. *Research in Language Testing*. Rowley, Mass.: Newbury House.

This volume is in seven parts. The first discusses the factors that make up language skill, the next four focus on proficiency and its measurement in each macroskill, the sixth considers the differences between native and non-native performance, and the last discusses measurement of factors that contribute to language learning. The only article specifically relevant to learning Japanese as a foreign language is one that considers the value of the MLAT aptitude test in predicting success in a university Japanese course. Only moderate correlations were found and other factors (e.g., experience in using the language with native speakers etc) were demonstrated to be also important.

Oskarsson, Mats. 1989. "Self-Assessment of Language Proficiency: Rationale and Applications". In *Language Testing*, Vol. 6, No. 1, 1989, pp. 1-11.

This paper provides a justification for self-assessment, indicates the benefits that flow to language learning from incorporating self-assessment in a language programme, and outlines some suitable techniques ranging from progress cards to rating scales to video and audio recordings.

Oskarsson, Mats. 1978. *Approaches to Self-Assessment in Foreign Language Learning*. Oxford: Pergamon.

This document was prepared in the context of the Council of Europe's work on language learning. It seeks to explore ways in which learners may be assisted to realize their own levels of achievement. It outlines some possible forms of self-assessment (both formal and informal), instances of where self-assessment is being developed and used, and sketches some research experiments.

Politzer, Robert L. and Mary McGroarty. 1983. "A Discrete Point Test of Communicative Competence". In *IRAL*, Vol. 21, No. 3, August 1983, pp. 179-191.

Politzer and McGroarty identify three definitions for communicative competence (ability to receive and convey information; appropriateness to topic, place and interlocutor; effective performance of speech acts). Three tests are devised to test each of these and the results compared with a grammar test. Linguistic competence is shown to be a condition for communicative competence but students can have the same level in the former and differ widely on the latter.

Raffaldini, Tina. 1988. "The Use of Situation Tests as Measures of Communicative Ability". In *Studies in Second Language Acquisition*, Vol. 10, No. 2, June 1988, pp. 197-216.

Distinguishes the oral proficiency interview and what it measures from communicative ability and proposes two situation tests as alternatives.

Read, John A. S. 1981. *Directions in Language Testing*. Singapore: RELC.

This collection of papers is from the 1980 RELC Seminar and includes papers on validating language tests, assessing oral proficiency, measuring communicative competence, cloze procedure, self-assessment, and the relationship between testing and teaching.

Rivera, Charlene. (ed.). 1984. *Language Proficiency and Academic Achievement*. Clevedon: Multilingual Matters.

These papers were presented at the Language Proficiency Symposium in Warrenton, Virginia, in 1981. They discuss theoretical frameworks for the concept of proficiency, social and cultural aspects of proficiency, and the need for research to clarify the concept.

Rivera, Charlene. (ed.). 1984a. *Communicative Competence Approaches to Language Proficiency Assessment: Research and Application*. Clevedon: Multilingual Matters.

These papers were presented at the Language Proficiency Assessment Symposium in March 1981 (presumably the same event as for the previous entry). The papers deal with definitions of terms (especially communicative competence), implications for testing, examples of proficiency assessment in different contexts, and psycholinguistic issues.

Savignon, Sandra J. 1985. "Evaluation of Communicative Competence: The ACTFL Provisional Proficiency Guidelines". In *The Modern Language Journal*, Vol. 69, No. 2, 1985, pp. 129-134.

This paper offers a critique of the ACTFL Guidelines, criticizing the inadequacy of the concept of "proficiency", preferring "communicative competence" but wanting a more adequate model, and asserting that they are unable to provide a basis for syllabus design.

Shohamy, Elana. 1982. "Predicting Speaking Proficiency from Cloze Tests: Theoretical and Practical Considerations for Test Substitution". In *Applied Linguistics*, Vol. 3, No. 3, pp. 161-171.

Shohamy compares results on cloze and oral interviews and concludes that it is better to look at ways of making the oral interview more efficient than to try to substitute other test types (specifically cloze) for it.

Skehan, Peter. 1988. "Language Testing. Part I", "State of the Art Article" in *Language Teaching*, Vol.21, No.4, pp. 211-221.

Skehan, Peter. 1989. "Language Testing. Part II", "State of the Art Article" in *Language Teaching*, Vol.22, No.1, pp. 1-13.

As the title indicates, this two-part article seeks to review current developments in language testing. It considers the nature of proficiency, communicative tests, ESP, achievement testing (including graded objectives), influences on test performance, statistical and analytic techniques, decision-making and test referencing, and public examinations.

Spolsky, Bernard. 1985. "The Limits of Authenticity in Language Testing". In *Language Testing*, Vol. 2, No. 1, 1985, pp. 31-40.

Spolsky argues for the importance of authenticity in test tasks since lack of authenticity weakens the generalizability of the results. All tests are inauthentic and only part of the problem is overcome with authentic-seeming tasks. Real-life observation is suggested as the only solution provided that the observers are sympathetic and trained.

Stansfield, Charles W. and Chip Harman (eds.). 1987. *ACTFL Proficiency Guidelines for the Less Commonly Taught Languages. A Familiarization Project for the Development of Proficiency Guidelines for Less Commonly Taught Languages*. New York, Washington D.C.: ACTFL/Center for Applied Linguistics.

This is a collection of papers on the use of the ACTFL Guidelines for teaching and assesment of "less commonly taught" languages. Papers include discussion of the OPI, the ILR interview, adaptation of the ACTFL Proficiency Guidelines to the less commonly taught languages, materials development, and a bibliography. There is specific discussion of the application of the ACTFL Guidelines to Arabic (a scale is provided), Indonesian, Hindi and African languages.

Stansfield, Charles W. and Dorry Mann Kenyon. 1988. *Development of the Portuguese Speaking Test. Year One Project Report. Development of Semi-Direct Tests of Oral Proficiency in Hausa, Hebrew, Indonesian and Portuguese*. Washington D.C.: Center for Applied Linguistics.

This publication reports the development of a SOPI in Portuguese. Interrater reliabilites were found to be uniform across both the taped and live tests. Most respondents preferred the live to the taped version.

Swain, Merrill. 1984. "Teaching and Testing Communicatively". In *TESOL Talk*, Vol. 15, Nos. 1/2, 1984, pp. 7-18.

This paper considers the general principles of communicative language testing and outlines a testing unit in a Grade 9 French course.

Trim, J. (ed.). 1988. *Evaluation and Testing in the Learning and Teaching of Languages for Communication*. Strasbourg: Council for Cultural Cooperation.

This report emanates from the Council of Europe's Project No. 12 "Learning and Teaching Modern Languages for Communication". Its papers deal with a number of issues relevant to communicative testing including assessing tests, personality as a factor in test development, graded objectives, educational evaluation, self-assessment and automatic data processing.

Underhill, Nic. 1987. *Testing Spoken Language: A Handbook of Oral Testing Techniques*. Cambridge: University Press.

This is a useful book giving practical guidance on writing oral tests suggesting test types, elicitation techniques, and marking systems and with one short chapter on basic concepts in test validity and reliability.

Valdman, Albert (ed.). 1988. *The Assessment of Foreign Language Oral Proficiency*. Special issue of *Studies in Second Language Acquisition*, Vol. 10, No. 2, June 1988.

This special issue contains articles on the ACTFL Scale and oral interview, the nature of proficiency, and other tests of oral proficiency.

Valdman, Albert. 1987. *Proceedings of the Symposium on the Evaluation of Foreign Language Proficiency*.

The book is in some nine sections dealing with the ACTFL Guidelines and their application to various languages and testing various aspects of language.

Valette, Rebecca M. 1967. *Modern Language Testing: A Handbook*. New York: Harcourt, Brace & World.

This book pre-dates developments in the concept and testing of proficiency and is included, like the books by Harris and Lado, as a standard reference in more traditional aspects of language testing. It discusses basic principles of test construction, administration and evaluation and has chapters on tests in the four macroskills and culture. An appendix lists with descriptions a number of commercially available tests in different languages (none in Japanese).

van Lier, Leo. 1989. "Reeling, Writhing, Drawling, Sketching and Fainting in Coils: Oral Proficiency Interviews as Conversation". In *TESOL Quarterly*, Vol. 23, No. 3, September 1989, pp. 489-508.

This paper provides a critique of the oral proficiency interview, looks at the differences between interviews and conversations, and discusses some of the major problems with interviews.

VanPatten, Bill. 1986. "The ACTFL Proficiency Guidelines: Implications for Grammatical Accuracy in the Classroom". In *Studies in Second Language Acquisition*, Vol. 8, No. 1, February 1986, pp. 56-67.

One of the implications of proficiency and proficiency scales seems to be to play down the importance of grammatical accuracy in favour of more global assessment and communicative skills. This paper reviews the arguments for a focus on grammatical accuracy from the beginning of a language learning programme in the light of language acquisition research and the ACTFL Proficiency Guidelines and concludes that arguments in favour of such a focus are misguided.

Walker, Carolyn R. 1983. "The Whole is More than the Sum of its Parts: Evaluation Criteria in Oral Tests". In *British Journal of Language Teaching*, Vol. 21, No. 1, pp. 41-44.

Walker considers the value of analytic as against global scales both from considerations of validity and reliability and from studies of the ARELS Oral EXamination. She concludes that there is no advantage in the use of more analytic scales and their use puts unreasonable demands on examiners.

Young, Dolly J. 1986. "The Relationship between Anxiety and Foreign Language Oral Proficiency Ratings". In *Foreign Language Annals*, Vol 19, No. 5, October 1986, pp. 439-445.

This paper examines the effect of anxiety on ACTFL OPI results. It is shown to have a lesser effect than ability.

C. BIBLIOGRAPHY

[NB. This bibliography does not list items already listed in the Annotated Bibliography.]

Adams, Raymond J., Patrick Griffin, and Lyn Martin. 1987. "A Latent Trait Method for Measuring a Dimension in Second Language Proficiency". In *Language Testing*, Vol. 4, No. 1, 1987, pp. 9-27.

Alderson, J. Charles. 1987. *Innovation in Language Testing: Can the Microcomputer Help? Language Testing Update*, Special Report No. 1. Lancaster: University of Lancaster.

Alderson, J. Charles and Arthur Hughes. 1981. *Issues in Language Testing*, ELT Documents No. 111. London: British Council.

Alderson, J. Charles and A. H. Urquhart. 1985. "The Effect of Students' Academic Discipline on their Performance on ESP Reading Tests". In *Language Testing*, Vol. 2, No. 2, 1985, pp. 192-204.

American Council on the Teaching of Foreign Languages. 1987 *Applications of Oral Proficiency Testing to Foreign Language Instruction*. New York: ACTFL.

- Antes, Richard L. 1985. "Teacher Education Admission Testing: Development and Implementation". (mimeograph) ED272498
- Babcock, Arthur. 1985. "Proficiency Testing in French at the University of Southern California". In *The French Review*, Vol. 58, No. 3, February 1985, pp. 403-410.
- Bachman, Lyle F.. 1988. "Problems in Examining the Validity of the ACTFL Oral Proficiency Interview". In *Studies in Second Language Acquisition*, Vol 10, No. 2, June 1988, pp. 149-164.
- Bachman, Lyle. 1986. "The Test of English as a Foreign Language as a Measure of Communicative Competence". In Stansfield 1986.
- Bachman, Lyle and Adrian S. Palmer. 1989. "The Construct Validation of Self-Ratings of Communicative Language Ability". In *Language Testing*, Vol. 6, No. 1, 1989, pp. 14-29.
- Bachman, Lyle and Sandra J. Savignon. 1986. "The Evaluation of Communicative Language Proficiency: A Critique of the ACTFL Oral Interview". In *Modern Language Journal*, Vol. 70, No. 4, Winter 1986, pp. 380-390.
- Baker, David. 1989. *Language Testing: A Critical Survey and Practical Guide*. London: Edward Arnold.
- Barnwell, David. 1987. "Oral Proficiency Testing in the United States". In *British Journal of Language Teaching*, Vol. 25, No. 1, Spring 1987, pp. 35-42.
- Briere, Eugene J. and Frances B. Hinofotis (eds.). 1979. *Concepts in Language Testing: Some Recent Studies*. Washington, D.C.: TESOL.
- Brindley, Geoff. 1986. *The Assessment of Second Language Proficiency: Issues and Approaches*. Adelaide: National Curriculum Resource Centre.
- Buck, Kathryn (ed.). 1989. *The ACTFL Oral Proficiency Interview Tester Training Manual*. Yonkers, NY: ACTFL.
- Buckby, Michael et al. 1981. *Graded Objectives and Tests for Modern Languages: An Evaluation*. London: Schools Council.
- Busch, D. 1982. "Introversion-Extroversion and the EFL Proficiency of Japanese Students". In *Language Learning*, Vol. 32, No. 1, 1982.
- Byram, Michael. 1985. "The Relationship between Graded Objectives and Testing in Foreign Language Teaching and Language Awareness Work". In Donmall, B. G. (ed.). *Language Awareness*. York: NCLE; 1985.
- Candlin, C. N. 1986. "Explaining Communicative Competence: Limits on Testability". In Stansfield 1986.
- Carroll, Brendan J. 1982. "The Implications for Testing of the Communicative Approach to Language Teaching". In *Studies in Language Learning*, No. 4, pp. 45-55.

- Carroll, Brendan J. and Patrick J. Hall. 1985. *Make your own Language Tests: A Practical Guide to Writing Language Performance Tests*. Oxford: Pergamon.
- Clark, John L. D. 1986. "A Study of the Comparability of Speaking Proficiency Interview Ratings across Three Government Language Training Agencies". Washington, D.C.: Center for Applied Linguistics.
- Clark, John L. D. 1983. "Language Testing: Past and Current Status - Directions for the Future". In *The Modern Language Journal*, Vol. 67, No. iv, 1983, pp. 431-441.
- Clark, John L. D. 1982. "Some Sociolinguistic and Discourse Analysis Considerations in Oral Proficiency Interviewing". Paper to a pre-conference session on "Analysis of Spoken Discourse", Georgetown University Roundtable on Languages and Linguistics, Washington D.C.. ED247762 FLO14530
- Clark, John L. D. 1972. *Foreign Language Testing: Theory to Practice*. Philadelphia, Pa.: Center for Curriculum Development.
- Clark, John L. D. and Li Ying-che. 1986. *Development, Validation, and Dissemination of a Proficiency-based Test of Speaking Ability in Chinese and an Associated Assessment Model for Other Less Commonly Taught Languages*. Washington, D.C.: Center for Applied Linguistics.
- Cohen, Andrew D. 1980. *Testing Language Ability in the Classroom*. Rowley, Mass.: Newbury House.
- Cohen, Andrew D. 1984. "On Taking Language Tests: What the Students Report". In *Language Testing*, Vol. 1, NO. 1, 1984, pp. 70-81.
- Cziko, Gary. 1984. "Some Problems with Empirically-based Models of Communicative Competence". In *Applied Linguistics*, Vol. 5, No. 1, 1984, pp. 23-37.
- Cziko, Gary A. 1982. "Improving the Psychometric, Criterion-Referenced, and Practical Qualities of Integrative Language Tests". In *TESOL Quarterly*, Vol. 16, No. 3, September 1982, pp. 367-379.
- Davies, Alan (ed.). 1968. *Language Testing Symposium: A Psycholinguistic Approach*. London: Oxford.
- Delamere, T. 1985. "Notional-Functional Syllabi and Criterion-Referenced Tests: The Missing Link". In *system*, Vol. 13, 1985.
- Douglas, Dan and Larry Selinker. 1985. "Principles for Language Tests within the 'Discourse Domains' Theory of Interlanguage: Research, Test Construction and Interpretation". In *Language Testing*, Vol. 2, No. 2, pp. 205-226.
- Educational Testing Service. 1986. *Oral Proficiency Familiarization Kit*. Boston, Ma.: Heinle & Heinle.
- Evans, C., J. Gaudin, and F. Raveau. 1987. "Testing Advanced Communicative Competence through Role-Play". In *British Journal of Language Teaching*, Vol. 25, No. 2, 1987.

- Foreign Service Institute. 1986. "Oral Language Proficiency Testing at the Foreign Service Institute". Washington, D.C.: FSI School of Language Studies.
- Green, P.S. and K. Hecht. 1987. "The Influence of Accuracy on Communicative Effectiveness". In *British Journal of Language Teaching*, Vol. 25, No. 2, 1987.
- Hale, G. A., Charles W. Stansfield, and R. P. Dwan. 1984. "A Comprehensive TOEFL Bibliography, 1963-82". In *The Modern Language Journal*, Vol. 68, No. 1, 1984.
- Harrison, Andrew. 1983. *A Language Testing Handbook*. London: Macmillan.
- Henning, Grant. 1987. *A Guide to Language Testing: Development, Evaluation, Research*. Rowley, Mass.: Newbury House.
- Henning, Grant, Thom Hudson and Jean Turner. 1985. "Item Response Theory and the Assumption of Unidimensionality for Language Tests". In *Language Testing*, Vol 2, No. 2, 1985, pp. 141-154.
- Hijirida, Kyoko. 1989. "ACTFL no nerai: Jituyo noryoku o ika ni sodateru ka" (Goals of ACTFL: How to Cultivate Practical Abilities). In *Nihongo*, June 1989.
- Hirsch, Bette G. 1985. "A Proficiency-Based French Conversation Course". In *The French Review*, Vol. LIX, No. 2, December 1985, p. 210 - 213.
- Hughes, Arthur and Don Porter (eds.). 1983. *Current Developments in Language Testing*. London: Academic Press.
- Hyltenstam, Kenneth and Manfred Pienemann. 1985. *Modelling and Assessing Second Language Acquisition*. Clevedon: Multilingual Matters.
- Ingram, D. E. 1990. "The International English Language Testing System (IELTS): Its Nature and Development". Paper to the 1990 RELC Seminar, Regional Language Centre, Singapore, 9-12 April, 1990. (mimeograph)
- Ingram, D. E. 1990a. "The International English Language Testing System (IELTS): The Speaking Test". Paper to the 1990 RELC Seminar, Regional Language Centre, Singapore, 9-12 April, 1990. (mimeograph)
- Ingram, D. E. 1985. "Assessing Proficiency: An Overview on Some Aspects of Testing". In Hyltenstam and Pienemann 1985: 215-276.
- Ingram, D. E. 1984. "Using Proficiency Rating Scales with High School Foreign Language Learners". Paper to the Biennial National Languages Conference of AFMLTA, Hobart, 31 August to 4 September, 1984. ED249767 FL014540
- Ingram, D. E. and Elaine Wylie. 1979/1985. *The Australian Second Language Proficiency Ratings*. (mimeograph). Also Canberra: DIEA/AGPS; 1984.
- Ingram, D. E. 1979/1982. "An Introduction to the Australian Second Language Proficiency Ratings". (mimeograph). Also Canberra: DIEA/AGPS; 1984.

- Jarvis, G. C. 1986. "Proficiency Testing: A Matter of False Hopes?" In *ADFL Bulletin*, Vol. 18, No. 1, 1986.
- Johnson, K. and D. Porter. 1983. *Perspectives in Communicative Language Teaching*. London: Academic Press.
- de Jong, John H.A.L. and C. A. W. Glas. 1987. "Validation of Listening Comprehension Tests using Item Response Theory". In *Language Testing*, Vol. 4, No. 2, December 1987, pp.170-194.
- Kaplan, Isabelle M. 1984. "Oral Proficiency Testing and the Language Curriculum: Two Experiments in Curricular Design for Conversation Courses". In *Foreign Language Annals*, Vol. 17, No. 5, 1984, pp. 491-497.
- Keitges, D. J. 1982. "Language Proficiency Interview Testing: An Overview". In *JALT Journal*, Vol. 4, 1982, pp. 17-45.
- Kramsch, Claire J. 1986. "Proficiency vs. Achievement: Reflections on the Proficiency Movement". In *ADFL Bulletin*, Vol. 18, No. 1, 1986, pp. 22-24.
- Krzanowski, Wojtek J. and Anthony J. Woods. 1984. "Statistical Aspects of Reliability in Language Testing". In *Language Testing*, Vol. 1, No. 1, 1984, pp. 3-19.
- Lantolf, James P. and Frawley, William. 1988. "Proficiency: Understanding the Construct". In *Studies in Second Language Acquisition*, Vol. 10, No. 2, 1988, pp. 181-195.
- Low, Graham D. 1982. "The Direct Testing of Academic Writing in a Second Language". In *System*, Vol. 10, No. 3, 1982, pp. 247-257.
- Lowe, Pardee. 1983. "The ILR Oral Interview: Origins, Applications, Pitfalls and Implications". In *Die Unterrichtspraxis*, Vol. 16, No. 2, 1983.
- Lowe, Pardee and J. E. Liskin-Gasparro. 1982. *Testing Speaking Proficiency: The Oral Interview*. Washington, D.C.: Center for Applied Linguistics.
- Lowe, Pardee and Charles Stansfield (eds.). 1988. *Second Language Proficiency Assessment: Current Issues*. Englewood Cliffs, NJ: Prentice Hall Regents.
- Lange, Dale L. and Pardee Lowe. 1988. "Rating Reading Passages according to the ACTFL Reading Proficiency Standard: Can it be Learned?". In *Foreign Language Annals*, Vol. 21, No. 3, May 1988, pp. 227-239.
- Madsen, Harold S. 1983. *Techniques in Testing*. Oxford: University Press.
- Magnan, Sally S. 1987. "Rater Reliability of the ACTFL Oral Proficiency Interview". In *Canadian Modern Language Review*, Vol. 43, No. 3, March 1987, pp. 525-537.

- Makino, Seeichi. 1987. "ACTFL gengo noryoku kijun to Amerika ni okeru nihongo kyoikuu" (ACTFL Proficiency Guidelines and Japanese Language Education in the U.S.). In *Nihongo Kyoiku*, No. 61, March, 1987.
- Massad, C. E. 1980. *Language Proficiency Tests Developed for the IEA International Study of Achievement in English as a Foreign Language*. ELT Document No. 108. London: British Council.
- McBride, John. 1990. "Computer Assisted Instruction in the Teaching of Japanese as a Second Language: An Issues Paper prepared for an Asialink Search Seminar held at the University of Wollongong, 24 May, 1990". (mimeograph)
- Morrow, Keith. 1990. "Evaluating Communicative Tests". Paper to the 1990 RELC Seminar, Regional Language Centre, Singapore, 9-12 April, 1990. (mimeograph)
- Morrow, Keith. 1985. "The Evaluation of Tests of Communicative Performance". In *Prospect*, Vol. 1, No. 2, 1985, pp. 32-44.
- Nunan, David. 1987. "The Ghost in the Machine: An Examination of the Concept of Proficiency". In *Prospect*, Vol. 2, No. 2, 1987.
- Oller, John W. 1979. *Language Tests at School*. London: Longman.
- Oller, John W. 1983. *Issues in Language Testing Research*. Rowley, Mass.: Newbury House.
- Omaggio, Alice C. 1985. *Proficiency, Curriculum, Articulation: The Ties that Bind*. Middlebury, VT.: North East Conference on the Teaching of Foreign Languages.
- Omaggio, Alice C. 1983. *Proficiency-Oriented Classroom Testing, Language in Education: Theory and Practice*, No. 52. Washington, D.C.: Center for Applied Linguistics.
- Palmer, Adrian S. 1981. "Testing Communication". In *IRAL*, Vol 10, 1981, pp. 35-45.
- Palmer, Adrian S. *The Construct Validation of Tests of Communicative Competence*. Washington, D.C.: TESOL.
- Palmer, Adrian S., Peter J. M. Groot, and George A. Trosper. 1981. *The Construct Validation of Tests of Communicative Competence*. Washington, D.C.: TESOL.
- Pienemann, Manfred and M. Johnston. 1986. "An Acquisition-Based Procedure for Second Language Assessment". In *Australian Review of Applied Linguistics*, Vol. 9, No. 1, 1986.
- Pienemann, Manfred, Malcolm Johnston and Geoff Brindley. 1988. "Constructing an Acquisition-based Procedure for Second Language Assessment". In *Studies in Second Language Acquisition*, Vol. 10, 1988, pp. 217-243.
- Piper, Terry H. and P. F. Leicester. 1980. "Word Association as an Indicator of English Language Proficiency". (mimeograph) ED227651

- Politzer, R. L. and M. McGroarty. 1983. "A Discrete Point Test of Communicative Competence". In *IRAL*, Vol 21, No. 3, 1983, pp. 179-191.
- Regents Examinations, Regents Competency Tests, and Proficiency Examinations: School Administrator's Manual*. New York: University of the State of New York, State Education Department and Bureau of Elementary and Secondary Testing Programs; 1989.
- Reisener, H. 1983. "Testing Communicative Ability". In *Practical English Teaching*, Vol. 3, No. 3, 1983.
- Reschke, C. 1984. "Oral Proficiency Testing: Meeting a Variety of Needs". In *Die Unterrichtspraxis*, Vol. 17, 1984, pp. 265-278.
- Rosser, Harry. 1986. "Language Proficiency Assessment: Implications and Applications". In *ADFL Bulletin*, Vol. 18, No. 1, 1986, pp. 15-19.
- Schulz, Renate A. 1988. "Proficiency-Based Foreign Language Requirements: A Plan for Action". In *ADFL Bulletin*, Vol. 19, No. 2, January 1988, pp. 24-28.
- Shohamy, Elana. 1983. "The Stability of Oral Proficiency Assessment on the Oral Interview Testing Procedures". In *Language Learning*, Vol. 33, No. 4, 1983, pp. 527-540.
- Skehan, Peter. 1983. "Issues in the Testing of ESP". In *MALS Journal*, Vol. 8, pp. 61-85.
- Skehan, Peter. 1984. "Issues in the Testing of English for Specific Purposes". In *Language Testing*, Vol. 1, No. 2, pp. 202-220.
- Spolsky, Bernard. 1989. "Communicative Competence, Language Proficiency, and Beyond". In *Applied Linguistics*, Vol 10, No. 2, pp. 138-156.
- Spolsky, Bernard. 1985. "What does it Mean to Know How to Use a Language? An Essay on the Theoretical Basis of Language Testing". In *Language Testing*, Vol. 2, No. 2, pp. 180-191.
- Stansfield, Charles W. 1989. *Simulated Oral Proficiency Interviews*. *ERIC Digest*, December 1989. Washington, D.C.: Center for Applied Linguistics.
- Stansfield, Charles W. (ed.). 1986. *Toward Communicative Competence Testing: Proceedings of the Second TOEFL Invitational Conference*. *TOEFL Research Reports No. 21*. Princeton, NJ: Educational Testing Service.
- Stansfield, Charles W. 1986a. *Technology and Language Testing*. Washington, D.C.: TESOL.
- Stansfield, Charles W. and Dorry Mann Kenyon. 1989. "A Report on Extending the ACTFL Proficiency Guidelines to the Less Commonly Taught Languages". In *Newsletter of the NorthEast Conference on the Teaching of Foreign Languages*, No. 26, pp. 28-30.

- Stevenson, Kiyo Yamada. 1989. "Opi shikenkan yosei no jittai repoto" (A report on OPI Tester Training). In *Nihongo*, June 1989.
- Sunderland, Jane et al. 1988. "To Be an Explainer: A Reporting Test of Oral Proficiency". In *ELT Journal*, Vol. 42, No. 1, January 1988, pp. 29-33.
- Takashima, Hideyuki. 1987. "To What Extent are Non-Native Speakers Qualified to Correct Free Composition? A Case Study". In *British Journal of Language Teaching*, Vol. 25, No. 1, 1987, pp. 43-48.
- Thompson, Irene, Richard T. Thompson and David Hipple. 1987. "Adapting the ACTFL/ETS Guidelines to the Less Commonly Taught Languages". In Stansfield and Harman 1987, Chapter 4.
- Thompson, Richard T. and Dora E. Johnson. 1988. *ERIC Digest: Proficiency, Testing in the Less Commonly Taught Languages*. Washington, D.C.: Center for Applied Linguistics.
- Usami, Mayumi. 1989. "OPI shikenkan yosei koza juko repoto" (A Trainee's Report on an OPI Tester Training Workshop). In *Nihongo*, June 1989.
- Wesche, Marjorie Bingham. 1983. "Communicative Testing in a Second Language". In *The Modern Language Journal*, Vol 67, No. 1, 1983, pp. 41-55.

D. RELEVANT TESTS, SCALES etc.

American Council of Teachers of Foreign Languages. 1982/1986. *ACTFL Proficiency Guidelines*. New York: ACTFL.

These scales were derived from the earlier published scales with clear influences from FSI, ILR and ASLPR. They provide a "hierarchy of global characterizations of integrated performance" in the four macroskills individually. They are intended to measure global performance using short descriptors for up to ten different levels in each macroskill. Versions are available for, amongst others, Chinese, Russian and Japanese as well as the generic English version.

Buros, Oscar K. (ed.). 1975. *Foreign Language Tests and Reviews*. Highland Park, NJ.: Gryphon Press.

Though the most recent edition seems to be 1975, this remains a useful listing of tests and reviews of them in areas relevant to language learning and teaching.

Educational Testing Service. 1981/1986. *Japanese Proficiency Test*. Princeton, N.J.: Educational Testing Service.

This norm-referenced test is used by many schools in the United States. It is intended for use with native English speakers learning Japanese as a foreign language with proficiencies ranging from that achieved after one year's study to that "associated with a college or graduate-level language program". It provides scores in listening and reading comprehension together with a diagnostic score for Japanese character recognition. The test is independent of any course and uses realistic language-use situations and contemporary texts. It uses fairly standard multiple-choice type questions to elicit comprehension levels. The results are a score between 20 and 80. A score-comparison table provides the distribution of scores for each of the three areas (listening comprehension, reading comprehension and character recognition) though no indication of the nature of the population (other than by "beginner", "intermediate" and "advanced") is provided. Similarly, there is no attempt made to provide a descriptive scale to facilitate interpretation of results.

The Japan Foundation. 1984/1989. *The Japanese Language Proficiency Test*. Tokyo: Japan Foundation/Association of International Education, Japan.

This norm-referenced test was developed by the Japan Foundation and the Association of International Education, Japan as the basis on which the proficiency of foreign students of Japanese may be certified. It is available once per year simultaneously in Japan and abroad. It is primarily intended for students wishing to go to Japan to study. The test has four levels ranging (according to the brief descriptive criteria provided) from Level 4 (which requires a knowledge of 100 words of kanji, 800 words of vocabulary, and the ability to engage in simple conversations) to Level 1 (where 2000 kanji and 10,000 words of vocabulary are expected together with the ability to survive in academic contexts). Each level has three sections: "writing-vocabulary", "hearing" and reading-grammar". "Passing scores" are set at 60% or 70% though no information is provided to indicate what this means in performance terms though Level 1 is necessary for entry to a Japanese university. The test is designed to be machine marked and is multiple-choice in format. The test seems to require a high level of "test-taking" expertise with long and quite complicated instructions as to how to fill in the answer sheets. However, it does not meet any of the criteria we have come to expect for a genuine proficiency test. In particular, the item format is not conducive to realistic language use and hence to the assessment of practical proficiency, it cannot be regarded as assessing productive skills, the results are, at best, difficult to interpret in terms of language behaviour, performance or practical proficiency and, as with any "indirect" test, one cannot readily identify the criteria on which such decisions are made as whether or not an applicant's proficiency is adequate for academic study in Japan. The information that accompanies the test provides extensive data on how many candidates from what countries took the test and their distribution of scores. Though this test has acquired considerable status, one cannot regard it as an adequate proficiency test and one would have to expect that it has the same shortcomings when used as such as do other "indirect" tests. (See also the entry under Munson and Harrison in the Annotated Bibliography above.)

Waites, Carol and Rena Kelly. 1989. "Towards TOLSKIP: A Review of Selected Language Testing Systems in relation to the Needs of the Australian Tourism Industry - Report to the Australia-Japan Foundation". (mimeograph).

Though some testers would wish to query certain of the statements made in this report for their accuracy of fact or interpretation, the report reviews a number of instruments in relation to their usefulness in meeting the needs of the tourism industry for measures of language proficiency. Though the assessments made and conclusions drawn are not always easy to defend, it is a useful list of instruments developed in Australia in recent years.